Available online at www.sciencedirect.com

SCIENCE DIRECT®

Discrete Mathematics 266 (2003) 293–309

DISCRETE MATHEMATICS

www.elsevier.com/locate/disc

A universal bound for a covering in regular posets and its application to pool testing☆

Vladimir I. Levenshtein

Keldysh Institute for Applied Mathematics, Russian Academy of Sciences, Miusskay Sq. 4, 125047 Moscow, Russia

Received 4 July 2001; received in revised form 5 April 2002; accepted 12 August 2002

Abstract

Let $V(n)$ be the set of all 2^n subsets of the set $N_n=\{1,2,\ldots,n\}$ and $V_i(n)=\{x\in V(n)\colon |x|=i\}$. For a fixed $i=1,\ldots,n$, consider a covering operator $F: V_i(n)\to V(n)$ such that $x\subseteq F(x)$ for any $x\in V_i(n)$. Let $C=\{F(x)\colon x\in V_i(n)\}$. For any $1\leqslant T\leqslant \binom{n}{i}$, consider the decreasing continuous function $g_i(T)=k+((k+1)/i)(1-\alpha)$ where k and α are uniquely defined by the conditions $T\binom{k}{i}=\alpha\binom{n}{i}$, $k\in\{i,\ldots,n\}$, and $1-i/(k+1)<\alpha\leqslant 1$. Using averaging and linear programing it is proved that

$$\frac{1}{\binom{n}{i}}\sum_{x\in V_i(n)}|F(x)|\geqslant g_i(|C|)\geqslant\frac{n}{\sqrt[i]{|C|}}$$

with the first inequality as an equality if and only if C is a Steiner $S(i,\{k,k+1\},n)$ design with a specified distance distribution. A generalization of this result to the case of monotone left-regular n-posets is given. One of motivating applications is the problem of reconstructing an unknown binary vector $\mathbf{x}$ of length n using pool testing under the condition that the vectors $\mathbf{x}$ are distributed with probabilities $p^{|x|}(1-p)^{n-|x|}$ where $x\in V(n)$ denotes the indices of the ones (active items) in $\mathbf{x}$. The bound above implies that the expected number of items which remain unresolved after application in parallel of arbitrary r pools is not less than

$$n\sum_{i=1}^{n}\binom{n}{i}p^i(1-p)^{n-i}2^{-(r/i)}-np.$$

This improves upon an information theoretic bound for the minimum average number $E(n,p)$ of tests to reconstruct an unknown $\mathbf{x}$ using two-stage pool testing and allows determination of

☆The research was partially supported by the Russian Foundation for Basic Research under grant 01-01-00035.

E-mail address: leven@keldysh.ru (V.I. Levenshtein).

doi:10.1016/S0012-365X(02)00815-4

the asymptotic behavior of $E(n,p)$ up to a positive constant factor as $n \to \infty$ under some restrictions upon $p = p(n)$.

Keywords: Covering; Poset; Pool testing; Steiner system; Universal bound

1. Introduction

We say that a finite graph $G=(V,E)$ is $(n+1)$-*partitioned* if there exists a partition of the vertex set V into $n+1$ enumerated disjoint subsets V_i, $V=\bigcup_{i=0}^{n} V_i$, such that (i) for every $x \in V_i$, $0 \leqslant i \leqslant n-1$, there exists $y \in V_{i+1}$ such that $(x,y) \in E$, (ii) for every $y \in V_i$, $1 \leqslant i \leqslant n$, there exists $x \in V_{i-1}$ such that $(x,y) \in E$, and (iii) for every $(x,y) \in E$, there exists i, $0 \leqslant i \leqslant n-1$, such that $x \in V_i$ and $y \in V_{i+1}$. For any $x \in V$ set $w(x)=i$ if $x \in V_i$ and call $w(x)$ the *weight* of x. Any $(n+1)$-partitioned graph $G=(V,E)$ can be treated as a partially ordered set (poset) $P=(V,\prec)$ with $x \prec y$ if and only if $w(x) \leqslant w(y)$ and there exists a path in G of length $w(y)-w(x)$ connecting x with y. This poset will also be denoted an *n-poset*. We now introduce some definitions and notations for an $(n+1)$-partitioned graph $G=(V,E)$ (and n-poset $P=(V,\prec)$). Throughout the paper we assume that $0 \leqslant i \leqslant j \leqslant n$. Let $v_i=|V_i|$, $L_{i,j}(x)=\{y\colon\ y \in V_j, x \prec y\}$ for all $x \in V_i$, and $L_{j,i}(y)=\{x\colon\ x \in V_i, x \prec y\}$ for all $y \in V_j$. Counting the total number of paths of length $j-i$ between V_i and V_j we have

$$\sum_{x \in V_i} |L_{i,j}(x)| = \sum_{y \in V_j} |L_{j,i}(y)|. \tag{1}$$

We call an $(n+1)$-partitioned graph or n-poset *right-regular* (*left-regular*) if, for all $0 \leqslant i \leqslant j \leqslant n$, $|L_{i,j}(x)|$ does not depend on $x \in V_i$ (respectively, $|L_{j,i}(y)|$ does not depend on $y \in V_j$). In these cases we denote $|L_{i,j}(x)|$ and $|L_{j,i}(y)|$ by $l_{i,j}$ and $l_{j,i}$, respectively. An n-poset is referred to as *regular* if it is right-regular and left-regular. For a regular n-poset (1) gives

$$v_i l_{i,j} = v_j l_{j,i} \quad \text{for any } 0 \leqslant i \leqslant j \leqslant n. \tag{2}$$

For a left-regular n-poset define

$$\beta_{i,j} = \frac{l_{j,i}}{l_{j+1,i}}, \quad 0 \leqslant i \leqslant j \leqslant n-1, \tag{3}$$

and call the n-poset *i-monotone* $(0 \leqslant i \leqslant n-1)$ if $\beta_{i,j}$ increases with j and $\beta_{i,n-1} < 1$. We call an n-poset *monotone* if it is *i-monotone* for all $i=1,\ldots,n-1$. The monotonicity of an n-poset means that for $i=1,\ldots,n-1$ the sequence $\{l_{j,i}\}$, $j=i,\ldots,n$, strictly increases and has the *concavity* property $l_{j-1,i}l_{j+1,i} < (l_{j,i})^2$, $j=i+1,\ldots,n-1$, i.e., the log-concavity relation $\ln l_{j,i} > \frac{1}{2}(\ln l_{j-1,i} + \ln l_{j+1,i})$ holds.

Example I (The subset poset). Let V be the set of all 2^n subsets of the set $N_n = \{1,2,\ldots,n\}$ and $V_i = \{x\colon\ x \in V, |x| = i\}$. Then $G=(V,E)$, with $V=\bigcup_{i=0}^{n} V_i$ and E consisting of all (x,y) such that y is obtained from x by addition of an element of N_n, forms a regular n-poset. In this case, $v_i = \binom{n}{i}$, $l_{i,j} = \binom{n-i}{j-i}$, and $l_{j,i} = \binom{j}{i}$. We call this poset the *subset poset*. For the subset poset, $\beta_{i,j} = 1 - i/(j+1)$, and it is a monotone n-poset.

Example II (Distance-regular posets). A graph $G=(V,E)$ of diameter n with the path distance $d(x,y)$ is called *distance-regular* [1,3,5] if there exist numbers a_i, b_i, c_i, such that, for any $x,y \in V$ with $d(x,y)=i$, the following holds:

$$|\{z\colon z\in V,\ d(x,z)=1,\ d(z,y)=i-1\}| = c_i, \quad i=1,\dots,n, \tag{4}$$

$$|\{z\colon z\in V,\ d(x,z)=1,\ d(z,y)=i+1\}| = b_i, \quad i=0,1,\dots,n-1, \tag{5}$$

$$|\{z\colon z\in V,\ d(x,z)=1,\ d(z,y)=i\}| = a_i, \quad i=0,1,\dots,n. \tag{6}$$

As a consequence the cardinality of any set $\{z\colon z\in V, d(x,z)=k, d(z,y)=l\}$ depends only on $d(x,y)$ and is denoted by $p^i_{k,l}$ if $d(x,y)=i$. The notion of distance-regular graph is equivalent to that of a metric association scheme with n classes [4]. Any vertex of a distance-regular graph of diameter n is at distance i from exactly v_i vertices where

$$v_0 = 1,\ v_i = \frac{b_0 b_1\cdots b_{i-1}}{c_2 c_3\cdots c_i} = v_{i-1}\frac{b_{i-1}}{c_i} \quad \text{for} \quad 1\leqslant i\leqslant n. \tag{7}$$

Therefore, denoting by V_i the set of vertices at distance i from a fixed vertex $x_0\in V$ and removing for any $x\in V_i$, $i=1,\dots,n$, all edges which connect x with a_i other vertices of V_i (if $a_i>0$), we get an $(n+1)$-partitioned graph. The corresponding poset is a regular n-poset with $l_{i,j}=p^i_{j,j-i}$ and $l_{j,i}=p^j_{i,j-i}$. Thus, all distance-regular graphs, in particular, the Hamming and Johnson graphs (metric spaces) H^n_q and J^n_w and their generalizations [4,5], give rise to regular n-posets which we call *distance-regular*. The distance-regular n-poset in the case of H^n_q has parameters $v_i=\binom{n}{i}(q-1)^i$, $l_{i,j}=\binom{n-i}{j-i}(q-1)^{j-i}$, and $l_{j,i}=\binom{j}{i}$ and this poset is monotone. In the case $q=2$ this poset coincides with the subset poset considered in Example I. It should be noted that the sequence (7) has the property $v_{i-1}v_{i+1}\leqslant (v_i)^2$ (see [1]). It follows that for the increasing sequence $s_i=\sum_{h=0}^{i} v_h$, $i=0,\dots,n$, the concavity condition $s_{i-1}s_{i+1}<(s_i)^2$ holds since it is equivalent to the inequality $s_{i-1}v_{i+1}<v_i s_i$ which is proved by induction using that

$$\frac{s_i}{s_{i+1}} = \frac{s_{i-1}+v_i}{s_i+v_{i+1}} < \frac{v_i}{v_{i+1}} \leqslant \frac{v_{i+1}}{v_{i+2}}.$$

This fact is useful to prove the monotonicity of some left-regular n-posets.

Example III (The subsequence poset). Consider the graph $G=(V,E)$ with $V=\bigcup_{i=0}^n V_i$ where $V_i=H^{n-i}_q$ (here H^l_q is the set of words of length l over an alphabet of q, $q\geqslant 2$, letters) and $(x,y)\in E$ if and only if one of x,y is obtained from another by deletion of one letter (this means that the deletion/insertion distance [11] between x and y is equal to one). This graph forms an n-poset with $v_i=q^{n-i}$ which will be called the *subsequence poset*. The subsequence poset is not right-regular. Nevertheless, as proved in [12] (see also [13]), it is left-regular with $l_{j,i}=\sum_{h=0}^{j-i}\binom{n-i}{h}(q-1)^h$. The subsequence poset is i-monotone for all i. This follows from the remark in the end of the previous example, since the sequence $l_{j,i}$, $j=i,\dots,n$, coincides with the increasing and concave sequence s_t, $t=0,\dots,n-i$ $(t=j-i)$, for the Hamming graph H^{n-i}_q.

For an n-poset $P=(V,\prec)$ and any $W\subseteq V$ consider a function $F: W\to V$ such that $x\prec F(x)$ for any $x\in W$. Such a function is called a *covering operator on W* and the set (code) $C(F,W)=\{F(x)\colon x\in W\}$ is called a *covering of W*. Note that this definition differs from the conventional one: Rather than taking a $y\in C$ to cover all $x\in V$ such that $x\prec y$, it covers only those x's for which $F(x)=y$. This investigation was stimulated by the following problem: For any probability distribution $\{p(x), x\in V\}$ and any covering operator F on V, find a lower bound to

$$\sum_{x\in V} p(x)(w(F(x))-w(x))$$

under the condition that the size of the covering $C(F,V)$ does not exceed a given number T. If the probability distribution is such that $p(x)=p(y)$ holds when $w(x)=w(y)$, this problem, in a certain sense, is reduced to the same problem for a covering of each set V_i with the uniform distribution on it.

For a covering operator $F: V_i\to V$ we estimate from below the value

$$\frac{1}{v_i}\sum_{x\in V_i} w(F(x)) \tag{8}$$

depending on the size of $C(F,V_i)$. A covering $C(F,V_i)$ is called a *Steiner i-design* in a poset $P=(V,\prec)$ if $x\prec y$, where $x\in V_i$ and $y\in C(F,V_i)$, implies that $y=F(x)$. (This means that each element of V_i precedes one and only one element of $C(F,V_i)$.) In Section 2 we introduce a continuous decreasing function $g_i(T)$ in real T, $v_i/l_{n,i}\leqslant T\leqslant v_i$, such that $g_i(v_i/l_{k,i})=k$. We prove that, for i-monotone left-regular posets, $g_i(|C(F,V_i)|)$ is a lower bound to (8) and find necessary and sufficient conditions for its tightness. These conditions state that C is a Steiner i-design formed by a specified number of elements of two successive weights. We also obtain a stronger lower bound when it is known that all elements of $C(F,V_i)$ have weight k or m. In Section 3 we calculate the general bounds of Section 2 for the subset poset and present some examples illustrating their attainability. Then we apply these bounds to pool testing and give a universal bound for the expected number of unresolved items for test matrices of a given size.

2. A universal bound for a covering in monotone left-regular posets

In this section we fix $i\in\{0,1,\dots,n\}$ and consider an i-monotone left-regular n-poset $P=(V,\prec)$. For a covering (code) $C=C(F,V_i)$ which corresponds to a covering operator F on V_i we estimate (8) from below using averaging and the linear programing method. This approach has previously been used by Knill [9]. Denote by $v_j(C)$ the components of the *weight distribution* of a code C, i.e., $v_j(C)=|V_j(C)|$ where $V_j(C)=V_j\cap C$, $j=i,\dots,n$. Since $l_{j,i}$ increases with j, $j\geqslant i$, for any covering C of V_i we have $|C|l_{n,i}\geqslant v_i\geqslant|C|$. First, we obtain a lower bound to (8) under some restrictions to the weight distribution of $C=C(F,V_i)$ and find necessary and sufficient conditions of its attainability. Then we get a universal lower bound to (8) as a specified function of $|C|$ alone. At last, we strengthen this bound in the case when the weight distribution of C consists of two non-successive components. In the next section we give examples which show that these bounds can be tight.

Theorem 1. *Let F be a covering operator on V_i and $C = C(F, V_i)$ be the corresponding covering of V_i for an i-monotone left-regular n-poset. Let for some k and m, $i \leqslant k < m \leqslant n$,*

$$\frac{v_i}{l_{m,i}} < |C| \leqslant \frac{v_i}{l_{k,i}} \tag{9}$$

with $v_j(C) = 0$ for all j such that $k < j < m$. Then

$$\frac{1}{v_i} \sum_{x \in V_i} w(F(x)) \geqslant k + (m-k) \frac{(v_i - |C| l_{k,i}) l_{m,i}}{v_i (l_{m,i} - l_{k,i})} \tag{10}$$

with equality if and only if C is a Steiner i-design,

$$v_k(C) = \frac{|C| l_{m,i} - v_i}{l_{m,i} - l_{k,i}}, \quad v_m(C) = \frac{v_i - |C| l_{k,i}}{l_{m,i} - l_{k,i}}, \tag{11}$$

and the remaining components of the weight distribution of C equal zero.

Proof. To simplify some notations we investigate

$$K(F, i, n) = \frac{1}{v_i} \sum_{x \in V_i} (w(F(x)) - i) = \frac{1}{v_i} \sum_{x \in V_i} w(F(x)) - i. \tag{12}$$

Set

$$J = J(i, k, m, n) = \{j\colon\ i \leqslant j \leqslant k \text{ or } m \leqslant j \leqslant n\}$$

and consider the linear programing problem (LPP) K to find

$$K(i, n) = \min \frac{1}{v_i} \sum_{x \in V_i} \sum_{y: w(y) \in J,\ x \prec y} (w(y) - i) u_{x,y} \tag{13}$$

where the minimum is taken over non-negative variables $u_{x,y}$, with

$$w(x) = i, \quad w(y) \in J \setminus \{i\}, \quad \text{and} \quad x \prec y$$

or

$$x = y \quad \text{and} \quad w(y) \in J,$$

satisfying the following linear inequalities:

$$u_{x,y} \leqslant u_{y,y} \quad \text{for any } x \prec y, \quad w(x) = i, \quad w(y) \in J \setminus \{i\}, \tag{14}$$

$$\sum_{y:\ w(y) \in J,\ x \prec y} u_{x,y} \geqslant 1 \quad \text{for any } x,\ w(x) = i, \tag{15}$$

$$\sum_{y:\ w(y) \in J} u_{y,y} \leqslant |C|. \tag{16}$$

Note that if these variables take values

$$u_{x,y} = \begin{cases} 1 & \text{if } F(x) = y \text{ or } x = y \in C, \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

then

$$\frac{1}{v_i} \sum_{x \in V_i} \sum_{y: w(y) \in J,\ x \prec y} u_{x,y}(w(y) - i) = K(F, i, n)$$

and all inequalities (14)–(16) hold, by (12) and the definition of a covering C of V_i. It follows that

$$K(F, i, n) \geqslant K(i, n). \tag{18}$$

Now we consider a simpler LPP L with $|J| - 1 = n - i - m + k + 1$ variables $u_{i,j},\ j \in J \setminus \{i\}$, and $|J| = n - i - m + k + 2$ variables $u_{j,j},\ j \in J$. The problem L consists of finding

$$L(i, n) = \min \sum_{j: j \in J} (j - i) \frac{v_j}{v_i}\, l_{j,i} u_{i,j}, \tag{19}$$

where the minimum is taken over these $2|J| - 1$ non-negative variables $u_{i,j}$ satisfying $|J| + 1$ linear inequalities

$$u_{i,j} \leqslant u_{j,j} \quad \text{for any } j \in J \setminus \{i\}, \tag{20}$$

$$\sum_{j: j \in J} v_j l_{j,i} u_{i,j} \geqslant v_i, \tag{21}$$

$$\sum_{j: j \in J} v_j u_{j,j} \leqslant |C|. \tag{22}$$

We next claim that $K(i, n) \geqslant L(i, n)$. It suffices to show that, for any feasible solution $\{u_{x,y}\}$ of LPP K, averaging

$$u_{i,j} = \frac{1}{v_j l_{j,i}} \sum_{x,y: x \prec y, w(x) = i, w(y) = j} u_{x,y} \quad \text{for any } j \in J \setminus \{i\} \tag{23}$$

and

$$u_{j,j} = \frac{1}{v_j} \sum_{y: w(y) = j} u_{y,y}, \quad \text{for any } j \in J \tag{24}$$

gives a feasible solution of L and the corresponding objective functions in (13) and (19) coincide. To prove this, we apply (23) and (24) in succession to (13)–(16). We get

$$\begin{aligned} \frac{1}{v_i} \sum_{x \in V_i} \sum_{y: x \prec y, w(y) \in J} (w(y) - i) u_{x,y} &= \frac{1}{v_i} \sum_{j \in J} \sum_{x,y: x \prec y, w(x) = i, w(y) = j} (j - i) u_{x,y} \\ &= \sum_{j \in J} (j - i) \frac{v_j}{v_i}\, l_{j,i} u_{i,j}; \end{aligned}$$

$$\sum_{x,y:x\prec y,w(x)=i,w(y)=j} u_{x,y} \leqslant l_{j,i} \sum_{y:w(y)=j} u_{y,y} = v_j l_{j,i} u_{j,j} \quad \text{for any } j \in J\backslash\{i\};$$

$$v_i \leqslant \sum_{j\in J} \frac{v_j l_{j,i}}{v_j l_{j,i}} \sum_{x,y:x\prec y,w(x)=i,w(y)=j} u_{x,y} = \sum_{j\in J} v_j l_{j,i} u_{i,j};$$

$$|C| \geqslant \sum_{j\in J} \sum_{y:w(y)=j} u_{y,y} = \sum_{j\in J} v_j u_{j,j}.$$

We have thus proved that

$$K(i,n) \geqslant L(i,n). \tag{25}$$

As in the case of LPPs of coding theory [14, Chapter 17] it is useful to consider the dual LPP to L for which $L(i,n)$ is the maximum of an objective function. Then the objective function for any feasible solution of the dual problem gives a lower bound to $L(i,n)$. Represent LPP L in the usual matrix form:

$$\text{minimize } \mathbf{u}\mathbf{d}^{\mathrm{t}} \text{ subject to } \mathbf{u} \geqslant \mathbf{0} \text{ and } \mathbf{u}A \leqslant -\mathbf{e}. \tag{26}$$

(Here $\mathbf{u}$, $\mathbf{d}$, $\mathbf{0}$, and $\mathbf{e}$ are real vectors, t denotes the transpose, and inequalities for vectors are treated as the corresponding inequalities for all their coordinates.) For LPP L we have

$$\mathbf{u} = (u_{i,i+1},\ldots,u_{i,k}u_{i,m},\ldots,u_{i,n},u_{i,i},\ldots,u_{k,k}u_{m,m},\ldots,u_{n,n}) \in R^{2|J|-1},$$

$$\mathbf{d} = (d_{i+1},\ldots,d_k,d_m,\ldots,d_n,0,\ldots,0) \in R^{2|J|-1} \quad \text{where } d_j = (j-i)\frac{v_j}{v_i} l_{j,i},$$

$$\mathbf{0} = (0,\ldots,0) \in R^{2|J|-1}, \quad \mathbf{e} = (1,0,\ldots,0,-|C|) \in R^{|J|+1},$$

and the $(2|J|-1)\times(|J|+1)$ matrix A is concatenation of a $(|J|-1)\times(|J|+1)$ matrix A_1 and a $|J|\times(|J|+1)$ matrix A_2. The matrix A_1 is obtained from the identity $(|J|-1)\times(|J|-1)$ matrix $E_{|J|-1}$ by adjoining a first column with entries $-(v_j/v_i)l_{j,i}$, $j\in J\backslash\{i\}$, and a final column of zeroes; A_2 is obtained from the matrix $-E_{|J|}$ by adjoining a last column with entries v_j, $j\in J$. Using linear programming duality (see, for example, [14]) we can express $L(i,n)$ with the help of the following LPP:

$$\text{maximize } \mathbf{e}\mathbf{w}^{\mathrm{t}} \text{ subject to } \mathbf{w} \geqslant \mathbf{0} \text{ and } A\mathbf{w}^{\mathrm{t}} \geqslant -\mathbf{d}^{\mathrm{t}}. \tag{27}$$

Introducing a vector $\mathbf{w} = (w_i,w_{i+1},\ldots,w_k,w_m,\ldots,w_n,\eta)$ of new variables we get that

$$L(i,n) = \max(w_i - \eta|C|), \tag{28}$$

where the maximum is taken over $|J|+1$ non-negative variables $w_i,w_{i+1},\ldots,w_k,w_m,\ldots,w_n,\eta$ satisfying $2|J|-1$ linear inequalities

$$\frac{w_j}{v_j} \geqslant \frac{l_{j,i}}{v_i}(w_i+i-j), \quad j\in J\backslash\{i\}, \tag{29}$$

$$\eta \geqslant \frac{w_j}{v_j}, \quad j \in J. \tag{30}$$

For any $\mu > 0$, we get a feasible solution of this problem setting

$$w_i = \mu, \quad w_j = \max\left((\mu + i - j)\frac{v_j}{v_i}\, l_{j,i}, 0 \right), \quad j \in J \setminus \{i\} \tag{31}$$

and

$$\eta = \frac{1}{v_i} \max_{j: j \in J} (\mu + i - j) l_{j,i}. \tag{32}$$

According to (18) and (25), for this feasible solution we have

$$K(F,i,n) \geqslant K(i,n) \geqslant L(i,n) \geqslant \mu - \eta |C|.$$

For a suitable selection of μ consider the function

$$f(j) = (\mu + i - j) l_{j,i} \tag{33}$$

and define μ by the condition $f(k) = f(m)$. It is easily seen that

$$\mu = k - i + \frac{(m-k) l_{m,i}}{l_{m,i} - l_{k,i}}. \tag{34}$$

Now we show that the maximum in (32) is attained at $j = k \in J$ and at $j = m \in J$. Using the inequalities

$$1 - (1-z)h \leqslant z^h < \frac{1}{1 + (1-z)h}$$

valid for any z, $0 < z < 1$, and any integer h, $h \geqslant 1$, and the fact that $\beta_j = \beta_{i,j}$ is less than 1 and increases with j, we have

$$\mu + i - k = \frac{(m-k) l_{m,i}}{l_{m,i} - l_{k,i}} = \frac{m-k}{1 - \beta_k \cdots \beta_{m-1}} \geqslant \frac{m-k}{1 - (\beta_k)^{m-k}} \geqslant \frac{1}{1 - \beta_k},$$

$$\mu + i - m = \frac{(m-k) l_{k,i}}{l_{m,i} - l_{k,i}} = \frac{m-k}{(\beta_k \cdots \beta_{m-1})^{-1} - 1} \leqslant \frac{m-k}{(\beta_{m-1})^{-(m-k)} - 1}$$
$$< \frac{1}{1 - \beta_{m-1}}.$$

Therefore, for $i \leqslant j < k$ we have

$$\begin{aligned} f(j+1) - f(j) &= (\mu + i - j - 1) l_{j+1,i} - (\mu + i - j) l_{j,i} \\ &= l_{j+1,i} \left((\mu + i - k + k - j)(1 - \beta_j) - 1 \right) \\ &\geqslant l_{j+1,i} \left(\frac{\beta_k - \beta_j}{1 - \beta_m} + (k-j)(1 - \beta_j) \right) > 0 \end{aligned}$$

and for $m \leqslant j < n$ we have

$$\begin{aligned} f(j+1) - f(j) &= (\mu + i - j - 1) l_{j+1,i} - (\mu + i - j) l_{j,i} \\ &= l_{j+1,i} \left((\mu + i - m + m - j)(1 - \beta_j) - 1 \right) \\ &< l_{j+1,i} \left(\frac{\beta_{m-1} - \beta_j}{1 - \beta_{m-1}} + (m - j)(1 - \beta_j) \right) < 0. \end{aligned}$$

Thus, the maximum in (32) is attained at $j = k \in J$ and at $j = m \in J$ (and only at these j). It follows that for the feasible solution (31) and (32) with μ given by (34) we get

$$\eta = (\mu + i - k) \frac{l_{k,i}}{v_i} = \frac{(m-k) l_{m,i} l_{k,i}}{v_i (l_{m,i} - l_{k,i})} \tag{35}$$

and

$$\mu - \eta |C| = k - i + \frac{(m-k) l_{m,i}}{l_{m,i} - l_{k,i}} - |C| \frac{(m-k) l_{m,i} l_{k,i}}{v_i (l_{m,i} - l_{k,i})}$$

which proves (10). If the bound (10) is attained for a covering $C = C(F, V_i)$ of V_i, then

$$K(F,i,n) = K(i,n) = L(i,n) = \mu - \eta |C|,$$

(23) and (24) with $u_{x,y}$ given by (17) form an optimal solution of the problem L, and (31) and (32) with μ given by (34) form an optimal solution of the dual problem. Note that for the solution of the dual problem we have $w_i = \mu > 0$ and $\eta > 0$ (see (34) and (35)). Moreover, equality in (30) is not possible for $j \in J$ different from k and m (for such $j \in J$, by (31), either $w_j = 0$ or $w_j = (v_j / v_i) f(j) < (v_j / v_i) f(k) = \eta v_j$). By the theorem of complementary slackness (see, for example, [14]), it implies that, for the optimal solution of the problem L, $u_{j,j} = 0$ for all $j \in J$, $j \neq k$, $j \neq m$, and hence

$$v_k(C) + v_m(C) = |C| \tag{36}$$

by (17) and (24). Set

$$r_j(C) = \sum_{y \in V_j(C)} \sum_{x: w(x) = i, F(x) = y} 1.$$

If for a covering $C = C(F, V_i)$ the bound (10) is attained, we have

$$\begin{aligned} & k r_k(C) + m r_m(C) = k v_i + (m-k) \frac{(v_i - |C| l_{k,i}) l_{m,i}}{l_{m,i} - l_{k,i}}, \\ & r_k(C) + r_m(C) = v_i. \end{aligned} \tag{37}$$

Therefore,

$$r_k(C) = l_{k,i} \frac{|C| l_{m,i} - v_i}{l_{m,i} - l_{k,i}}, \quad r_m(C) = l_{m,i} \frac{v_i - |C| l_{k,i}}{l_{m,i} - l_{k,i}} \tag{38}$$

and hence

$$\frac{r_k(C)}{l_{k,i}} + \frac{r_m(C)}{l_{m,i}} = |C|.$$

Note that for $j=k$ and $j=m$ we have $r_j(C) \leqslant l_{j,i}v_j(C)$ with equality if and only if for any $y \in V_j(C)$ there exist exactly $l_{j,i}$ elements $x \in V_i$ such that $F(x)=y$. Taking account of (36) we conclude that $r_k(C)=l_{k,i}v_k(C)$, $r_m(C)=l_{m,i}v_m(C)$, and C is a Steiner i-design. This proves the necessity of the conditions for attainability of (10). The sufficiency of these conditions easily follows from (11). □

Note that the statement of Theorem 1 is evident if $|C|l_{k,i}=v_i$, $k=i,\dots,n$. In this case $v_m(C)=0$ and the right-hand side of (10) equals k.

Our next goal, for any i, $i=0,1,\dots,n$, and any i-monotone left-regular n-poset, is to introduce a decreasing continuous function $g_i(T)$ in real T such that, for any covering $C=C(F,V_i)$, $g_i(|C|)$ is a lower bound to (8). By i-monotonicity of the n-poset, for any real T, $v_i/l_{i,n} < T \leqslant v_i$, there exists a unique integer $k=k(T)$, $k=i,\dots,n-1$, such that

$$\frac{v_i}{l_{k+1,i}} < T \leqslant \frac{v_i}{l_{k,i}}. \tag{39}$$

We put $k(T)=n$ when $T=v_i/l_{n,i}$ and consider the function

$$g_i(T)=k+\frac{v_i-Tl_{k,i}}{v_i(1-\beta_{i,k})} \text{ with } k=k(T). \tag{40}$$

Note that this function is decreasing and continuous for $v_i/l_{i,n} \leqslant T \leqslant v_i$, since the second member in the sum (40) equals 1 for $T=v_i/l_{k+1,i}$. Applying Theorem 1 for $k=k(|C|)$ and $m=k(|C|)+1$ we get the following statement.

Theorem 2. *Let F be a covering operator on V_i and $C=C(F,V_i)$ be the corresponding covering of V_i for an i-monotone left-regular n-poset. Then*

$$\frac{1}{v_i}\sum_{x\in V_i} w(F(x)) \geqslant g_i(|C|) \tag{41}$$

with equality if and only if C is a Steiner i-design,

$$v_k(C)=\frac{|C|l_{k+1,i}-v_i}{l_{k+1,i}-l_{k,i}}, \quad v_{k+1}(C)=\frac{v_i-|C|l_{k,i}}{l_{k+1,i}-l_{k,i}}, \text{ where } k=k(|C|), \tag{42}$$

and the remaining components of the weight distribution of C equal zero.

We also consider the case when only two components of the weight distribution of $C=C(F,V_i)$, say $v_k(C)$ and $v_m(C)$, are positive. Condition (9) suffices for this case because of the i-monotonicity of the n-poset.

Theorem 3. *Let F be a covering operator on V_i and $C=C(F,V_i)$ be the corresponding covering of V_i for an i-monotone left-regular n-poset. If $v_k(C)>0$ and $v_m(C)\geqslant 0$ for some k and m, $i\leqslant k<m\leqslant n$, and the remaining components of the weight distribution of C equal zero, then*

$$\frac{1}{v_i}\sum_{x\in V_i} w(F(x)) \geqslant k+(m-k)\frac{(v_i-|C|l_{k,i})l_{m,i}}{v_i(l_{m,i}-l_{k,i})} \tag{43}$$

with equality if and only if C is a Steiner i-design,

$$v_k(C) = \frac{|C| l_{m,i} - v_i}{l_{m,i} - l_{k,i}}, \quad v_m(C) = \frac{v_i - |C| l_{k,i}}{l_{m,i} - l_{k,i}}. \tag{44}$$

3. A universal bound for a covering in the subset poset and its application

In this section we study a covering in the subset poset mentioned in Example I. The subset poset $(V, \prec)$ consists of the set $V = V(n)$ of all subsets of an n-set ordered by inclusion. In this case the weight $w(x)$ equals the cardinality of the subset x, $v_i = \binom{n}{i}$, $l_{i,j} = \binom{n-i}{j-i}$, $l_{j,i} = \binom{j}{i}$. This regular poset is monotone since $\beta_{i,j} = \binom{j}{i}/\binom{j+1}{i} = 1 - (i/(j+1))$ is an increasing function in j (for $i \geqslant 1$). The Steiner i-designs C in the subset poset are the classical Steiner $S(i, K, n)$ designs where $K = \{j\colon v_j(C) > 0\}$.

The subset poset is the main object for covering and packing problems for sets V_i (see, for example, [4]). However, it should be noted that, by our definition, $y \in C = C(F, V_i)$ does not “cover” $x \in V_i$ if $x \subseteq y$ but $F(x) \neq y$. Distinctness of the problem under consideration from other well-known problems also consists of the fact that we minimize the average value (8) for a given cardinality of a covering $C = C(F, V_i)$. Investigation of this problem was stimulated by applications in pool testing problems which will be considered in the end of the section.

Corollary 1. *Let F be a covering operator on V_i, $i \geqslant 1$, and $C = C(F, V_i)$ be the corresponding covering of V_i in the subset n-poset. If k is defined by the condition*

$$\frac{\binom{n}{i}}{\binom{k+1}{i}} < |C| \leqslant \frac{\binom{n}{i}}{\binom{k}{i}},$$

then

$$\frac{1}{\binom{n}{i}} \sum_{x \in V_i} w(F(x)) \geqslant g_i(|C|) = k + \frac{k+1}{i} \left(1 - |C| \frac{\binom{k}{i}}{\binom{n}{i}} \right) \tag{45}$$

with equality if and only if C is a Steiner $S(i, \{k, k+1\}, n)$ design,

$$v_k(C) = \frac{k+1}{i} |C| - \frac{\binom{n}{i}}{\binom{k}{i-1}}, \quad v_{k+1}(C) = \frac{\binom{n}{i}}{\binom{k}{i-1}} - \frac{k+1-i}{i} |C|. \tag{46}$$

Corollary 2. *Let F be a covering operator on V_i, $i \geqslant 1$, and $C = C(F, V_i)$ be the corresponding covering of V_i for the subset n-poset. If $v_k(C) > 0$ and $v_m(C) \geqslant 0$ for some k and m, $i \leqslant k < m \leqslant n$, and the remaining components of the weight distribution of C equal zero, then*

$$\frac{1}{\binom{n}{i}} \sum_{x \in V_i} w(F(x)) \geqslant k + (m-k) \frac{\binom{m}{i} \left(\binom{n}{i} - |C| \binom{k}{i} \right)}{\binom{n}{i} \left(\binom{m}{i} - \binom{k}{i} \right)} \tag{47}$$

with equality if and only if C is a Steiner $S(i,\{k,m\},n)$ design,

$$v_k(C)=\frac{|C|\binom{m}{i}-\binom{n}{i}}{\binom{m}{i}-\binom{k}{i}},\quad v_m(C)=\frac{\binom{n}{i}-|C|\binom{k}{i}}{\binom{m}{i}-\binom{k}{i}}. \tag{48}$$

As examples apply Corollaries 1 and 2 for $i=4$. Let a covering $C=C(F,V_4)$ exist such that $|C|=\frac{n+10}{60}\binom{n-1}{3}$. Then $k=k(|C|)=5$ (for $n\geqslant 5$) and (45) shows that

$$\frac{1}{\binom{n}{4}}\sum_{x\in V_4} w(F(x))\geqslant 6-\frac{5}{n}.$$

This bound is attained if and only if C is a Steiner $S(4,\{5,6\},n)$ design, with $v_5(C)=\frac{1}{4}\binom{n-1}{3}$ and $v_6(C)=\frac{n-5}{60}\binom{n-1}{3}$. For $n=17$ such a design was found by Kramer and Mathon in [10] and then they were constructed by Tonchev [15] for all $n=4^l+1$, $l=2,3,\dots$, using the Preparata codes of length 4^l. In this construction 6-subsets which correspond to the minimum weight codewords are used as $V_6(C)$ and all 4-sets which are not covered by these codewords are extended by an additional element (∞) and used as $V_5(C)$. (Here it is essential that these 4-sets form a Steiner $S(3,4,4^l)$ design as it was proved by Zinoviev in [16].) Applying Corollary 1 to a covering $C=C(F,V_4)$ such that $|C|=\frac{n+11}{60}\binom{n}{3}$ we get that $k=k(|C|)=5$ (for $n\geqslant 10$) and

$$\frac{1}{\binom{n}{4}}\sum_{x\in V_4} w(F(x))\geqslant 6-\frac{7}{n-3}.$$

This bound can be improved by Corollary 2 if it is known that the weight distribution of C has only two non-zero components $v_4(C)$ and $v_6(C)$. In this case (47) shows that

$$\frac{1}{\binom{n}{4}}\sum_{x\in V_4} w(F(x))\geqslant 6-\frac{2}{n-3}.$$

This bound is also valid under the weaker restriction that $v_5(C)=0$, by Theorem 1. It is attained if and only if C is a Steiner $S(4,\{4,6\},n)$ design with $v_4(C)=\frac{1}{4}\binom{n}{3}$ and $v_6(C)=\frac{n-4}{60}\binom{n}{3}$. Such $S(4,\{4,6\},n)$ designs exist for $n=4^l$, $l=2,3,\dots$. They are formed by 6-sets which correspond to the minimum weight words of the Preparata code and by 4-sets which do not belong to these 6-sets.

Corollary 3. *Let F be a covering operator on V_i, $i\geqslant 1$, and $C=C(F,V_i)$ be the corresponding covering of V_i for the subset n-poset. Then*

$$\frac{1}{\binom{n}{i}}\sum_{x\in V_i} w(F(x))\geqslant \frac{n}{\sqrt[i]{|C|}}. \tag{49}$$

Proof. Note that $g_i(T)=k+((k+1)/i)(1-\alpha)$ where k and α are uniquely defined by the conditions

$$T=\alpha\frac{\binom{n}{i}}{\binom{k}{i}}\quad\text{and}\quad \beta_{i,k}=1-\frac{i}{k+1}<\alpha\leqslant 1.$$

Therefore, by Corollary 1, it suffices to show that for any fixed i and k, $i \leqslant k \leqslant n$, we have

$$h(\alpha) = k + \frac{k+1}{i}(1-\alpha) - \frac{n}{\sqrt[i]{\alpha \frac{\binom{n}{i}}{\binom{k}{i}}}} \geqslant 0 \quad \text{if} \quad 1 - \frac{i}{k+1} \leqslant \alpha \leqslant 1.$$

This is true because the second derivative of $h(\alpha)$ is negative for $0 < \alpha \leqslant 1$,

$$h(1) = k - n \left(\frac{\binom{n}{i}}{\binom{k}{i}} \right)^{-1/i} \geqslant 0,$$

and

$$h\left(1 - \tfrac{i}{k+1}\right) = k + 1 - n \left(\frac{\binom{n}{i}}{\binom{k+1}{i}} \right)^{-1/i} \geqslant 0.$$

(Here we used $\binom{n}{i}/\binom{j}{i} \geqslant \left(\frac{n}{j}\right)^i$ if $i \leqslant j \leqslant n$.) $\square$

Apply Corollary 3 to obtain a universal bound for a covering operator F on the entire $V(n)$ and the corresponding code $C = C(F, V(n))$. We assume there is a given probability distribution $p(x)$ on $V(n)$, $x \in V(n)$. A probability distribution is referred to as *symmetric*, if $p(x) = p(y)$ when $|x| = |y|$. The Bernoulli p-scheme ($0 < p < 1$) with $p(x) = p^{|x|}(1-p)^{n-|x|}$ is an example of a symmetric distribution.

Corollary 4. *Let F be a covering operator on $V(n)$ with a symmetric probability distribution $p(x)$ and $C = C(F, V(n))$ be the corresponding covering of $V(n)$. Then*

$$\sum_{x \in V(n)} p(x)|F(x)| \geqslant \sum_{i=1}^{n} \frac{n}{\sqrt[i]{|C|}} \sum_{x \in V_i} p(x) \tag{50}$$

and, in particular, in the case of the Bernoulli p-scheme,

$$\sum_{x \in V(n)} p(x)|F(x)| \geqslant n \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} (|C|)^{-1/i}. \tag{51}$$

Proof. Let $C_i = \{F(x) \colon x \in V_i\}$ and $P(i) = \sum_{x \in V_i} p(x)$. Using $|C_i| \leqslant |C|$ and $p(x) = 1/\binom{n}{i} P(i)$ if $x \in V_i$ we have

$$\sum_{x \in V(n)} p(x)|F(x)| = \sum_{i=0}^{n} \frac{P(i)}{\binom{n}{i}} \sum_{x \in V_i} |F(x)| \geqslant \sum_{i=1}^{n} \frac{nP(i)}{\sqrt[i]{|C_i|}} \geqslant \sum_{i=1}^{n} \frac{n}{\sqrt[i]{|C|}} \sum_{x \in V_i} p(x),$$

by Corollary 3. $\square$

Now we apply Corollary 4 to the problem of reconstructing an unknown vector $\mathbf{x} \in \{0,1\}^n$ using the pool testing procedure [8,7]. There exists a one-to-one mapping between vectors $\mathbf{x}=(\xi_1,\dots,\xi_n)\in\{0,1\}^n$ and subsets $x=\{i\colon\ \xi_i=1\}\in V(n)$ and we use for them the same letter but different fonts. For a vector $\mathbf{x}\in\{0,1\}^n$, we call elements of x and $\{1,2,\dots,n\}\backslash x$ respectively by *active* and *inactive* items of $\mathbf{x}$. Consider a binary $r\times n$ matrix H with rows $\mathbf{h}_i=(h_{i,1},\dots,h_{i,1})\in\{0,1\}^n$, $i=1,\dots,r$, and define values

$$s_i=\bigvee_{j=1}^{n}\xi_j\&h_{i,j},\qquad i=1,\dots,r, \tag{52}$$

where $\vee$ and & denote the logical operations of disjunction and conjunction. The value s_i can be treated as a result of application of a *test* given by the vector $\mathbf{h}_i$ to an unknown vector $x\in V(n)$. Note that (52) can be rewritten as

$$s_i=\begin{cases}1 & \text{if } x\cup h_i \text{ is not empty,}\\ 0 & \text{otherwise,}\end{cases} \tag{53}$$

i.e., $s_i=1$ if and only if the set h_i of active items of the test vector $\mathbf{h}_i$ (called a *pool*) has non-empty intersection with the set x of active items of $\mathbf{x}$. The problem is to reconstruct an unknown $\mathbf{x}$ solving a system (52) of r logical equations with n boolean variables.

Given a test $r\times n$ matrix H, the system (52) in general does not have a unique solution. Denote by $\mathbf{s}(x)$ the vector $\mathbf{s}=(s_1,\dots,s_r)\in\{0,1\}^r$, where s_i are defined by (52), and call it a *syndrome* of $\mathbf{x}=(\xi_1,\dots,\xi_n)\in\{0,1,\}^n$ (and of the corresponding $x\in V(n)$). For a given vector $\mathbf{s}=(s_1,\dots,s_r)\in\{0,1\}^r$, the (possibly empty) set $Q(H,\mathbf{s})$ of all solutions of the system (52) can be expressed as

$$Q(H,\mathbf{s})=\{\mathbf{x}\colon\ \mathbf{x}\in\{0,1\}^n,\ \ \mathbf{s}(x)=\mathbf{s}\}.$$

Set

$$X(\mathbf{s})=\bigcup_{x\colon\ x\in V(n),\ \mathbf{x}\in Q(H,\mathbf{s})} x \tag{54}$$

and note that if $Q(H,\mathbf{s})$ is not empty, then it contains the vector $\mathbf{X}(\mathbf{s})\in\{0,1\}^n$ having the set of active items $X(\mathbf{s})\in V(n)$, by (52) and the definition of disjunction.

Although in general we cannot uniquely recover $\mathbf{x}=(\xi_1,\dots,\xi_n)$ from its syndrome $\mathbf{s}=\mathbf{s}(x)$, we may be able to determine some of its components. For a given H and $\mathbf{s}$ we call an item $j\in\{1,2,\dots,n\}$ *positive* or *negative*, respectively, if the jth component of all vectors of $Q(H,\mathbf{s})$ is 1 (active) or 0 (inactive), respectively. The remaining items $i\in I_n$ we call *unresolved*. For any $x\in\{0,1\}^n$ we denote by $u(H,\mathbf{x})$ the number of unresolved items when $\mathbf{s}=\mathbf{s}(x)$. How is it possible to determine from the matrix H and syndrome $\mathbf{s}$ what items are negative, positive and unresolved? It is easily seen that an item j is negative if and only if there exists a pool h_i such that $j\in h_i$ (or $h_{i,j}=1$) and $s_i=0$. If for an item j there exists a pool h_i such that $s_i=1$ and h_i contains j and all other of its active items (if they exist) are negative, then this item is positive. (From the next statement it follows that this condition is also necessary for an item j to be positive.) In the remaining cases either all pools do not contain j or every pool h_i, such that $s_i=1$ and h_i contains j, contains also at least one more item which is not negative. In this case the item j is unresolved since $G(H,\mathbf{s})$ contains $X(\mathbf{s})$ and the vector obtained from it by replacing its jth component 1 by 0. As an example consider the following 4×6 test matrix H and the syndrome $\mathbf{s} = (1011)$:

x_1	x_2	x_3	x_4	x_5	x_6	$\mathbf{s}$
1	1	1	0	0	0	1
1	0	0	1	1	0	0
0	1	0	0	1	1	1
0	0	0	1	1	1	1

In this case items 1, 4, and 5 are negative, item 6 is positive, and items 2 and 3 are unresolved. We also have $Q(H,\mathbf{s}) = \{(010001), (001001), (011001)\}$, $\mathbf{X}(\mathbf{s}) = (011001)$, and $X(\mathbf{s}) = \{2, 3, 6\}$.

Corollary 4 can be used to estimate the average number of unresolved items. We identify the probability of a vector $\mathbf{x} \in \{0,1\}^n$ with the probability $p(x)$ of the set $x \in V(n)$ of its active items.

Theorem 4. *For a symmetric probability distribution $p(x)$, $x \in V(n)$, and any $r \times n$ matrix H,*

$$\sum_{x \in V(n)} p(x)u(H,\mathbf{x}) \geqslant \sum_{i=1}^{n} n2^{-r/i} \sum_{x \in V_i} p(x) - \sum_{i=0}^{n} \sum_{x \in V_i} i\,p(x) \tag{55}$$

and, in particular, in the case of the Bernoulli p-scheme,

$$\sum_{x \in V(n)} p(x)u(H,\mathbf{x}) \geqslant n \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} 2^{-r/i} - np. \tag{56}$$

Proof. Consider the operator $F : V(n) \to V(n)$ defined as follows:

$$F(x) = X(\mathbf{s}(x)). \tag{57}$$

By (54), $x \subseteq X(\mathbf{s}(x))$, and hence F is a covering operator on $V(n)$. Note that each item $j \in X(\mathbf{s}(x)) \backslash x$ is unresolved, because both $x \in V(n)$ and $X(\mathbf{s}(x)) \in V(n)$ are solutions of the system (52). It follows that

$$u(H,\mathbf{x}) \geqslant |X(\mathbf{s}(x)) \backslash x| = |F(x)| - |x|,$$

and Theorem 4 follows from Corollary 4. $\square$

A *two-stage testing* to reconstruct an unknown $\mathbf{x} = (\xi_1, \ldots, \xi_n) \in \{0,1\}^n$ consists of calculating the syndrome $\mathbf{s} = (s_1, \ldots, s_r)$, where s_i are defined by (52), in Stage 1 and then conducting individual tests (pools of cardinality 1) in Stage 2 (tests corresponding to appending to H rows each of which has only one active item) in order to determine which of the $u(H,\mathbf{x})$ items that are unresolved after Stage 1 actually are active and which are inactive. Assuming that the choice of $\mathbf{x} \in \{0,1\}^n$ is governed by a Bernoulli p-scheme we characterize the efficiency of this two-stage testing by the average number

$$E(H,p) = r + \sum_{\mathbf{x} \in \{0,1\}^n} p(\mathbf{x})u(H,\mathbf{x}) \tag{58}$$

of tests used to determine an unknown $\mathbf{x} \in \{0,1\}^n$. The problem is to find $E(n,p) = \min E(H,p)$ where the minimum is taken over all test matrices H with n columns (and any number $r \geqslant 1$ of rows). Applications of two-stage testing abound; see, for example, [7,9].

The Shannon theorem on the average length of a prefix code implies (see [2]) the following information theoretic bound:

$$E(n,p) \geqslant n\left(p\log_2\frac{1}{p} + (1-p)\log_2\frac{1}{1-p}\right). \tag{59}$$

This bound on the average number of tests is valid for any *adaptive* testing algorithm, not just two-stage algorithms. Eq. (59) implies $E(n,p) = o(n)$ as $n \to \infty$ only if $p \to 0$. Another universal (lower) bound is obtained if we estimate (58) using Theorem 4 and then minimize it over $r \geqslant 1$. The investigation of this bound (see [2]) shows that it is asymptotically better than the information theoretic bound (59) as $n \to \infty$ when $p \leqslant c(\ln n/n)$ with any constant $c > 0$. Moreover, together with a certain random selection bound, it determines the asymptotic behavior of $E(n,p)$ up to a positive constant factor when p is not too small (namely $p > n^{2-\varepsilon}$ with $\varepsilon > 0$ as small as one wishes).

In conclusion it is worth to note that the covering operator F on $V(n)$ defined by (57) has the additional properties: $F(F(x)) = F(x)$ for any $x \in V(n)$ and $F(x) \subseteq F(y)$ if $x \subseteq y$, i.e., it is a *closure* operator on $V(n)$. An interesting open problem is to improve upon Corollary 4 and Theorem 4 in the class of closure operators.

References

[1] E. Bannai, T. Ito, Algebraic Combinatorics. I. Association Schemes, Benjamin/Cummings, London, 1984.

[2] T. Berger, V.I. Levenshtein, Asymptotic efficiency of two-stage disjunctive testing, IEEE Trans. Inform. Theory 48 (7) (2002) 1741–1749.

[3] A.E. Brouwer, A.M. Cohen, A. Neumaier, Distance-Regular Graphs, Springer, Berlin, 1989.

[4] P. Delsarte, An algebraic approach to the association schemes of coding theory, Philips Res. Rep. (Suppl.) 10 (1973).

[5] P. Delsarte, Association schemes and t-design in regular semilattices, J. Combin. Theory Ser. A 20 (1976) 230–273.

[6] J.H. Dinitz, D.R. Stinson (Eds.), Contemporary Design Theory, Wiley Interscience Publication, New York, 1992.

[7] D.-Z. Du, F.K. Hwang, Combinatorial Group Testing and its Applications, World Scientific, Singapore, 1993.

[8] W.H. Kautz, R.R. Singleton, Nonrandom binary superposed codes, IEEE Trans. Inform. Theory 10 (4) (1964) 363–377.

[9] E. Knill, Lower bounds for identifying subset members with subset queries. Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, CA, January 22–24, 1995, 369–377 (Chapter 40).

[10] E.S. Kramer, R. Mathon, Proper $S(t,K,v)$'s for $t \geqslant 3$, $v \leqslant 16$, $|K| > 1$ and their extensions, J. Combin. Designs 3 (1995) 411–426.

[11] V.I. Levenshtein, Binary codes capable of correcting deletions, insertions and reversals, Dokl. Akad. Nauk. 163(4) (1965) 845–848 (in Russian). English translation in Sov. Phys. - Dokl. 10 (8) (1966) 707–710.

[12] V.I. Levenshtein, Elements of coding theory, in: Discrete Mathematics and Math. Problems of Cybernetics, Nauka, Moscow, 1974, pp. 207–305 (in Russian). Diskrete Mathematik und Mathematische Fragen der Kybernetik, Academie-Verlag, Berlin, 1980, pp. 198–279 (German translation).

[13] V.I. Levenshtein, Efficient reconstruction of sequences from their subsequences or supersequences, J. Combin. Theory Ser. A 93 (2) (2001) 310–332.

[14] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland, Amsterdam, 1977.

[15] V.D. Tonchev, A class of Steiner 4-wise balanced designs derived from Preparata codes, J. Combin. Designs 4 (1996) 203–204.

[16] V.A. Zinoviev, Combinatorial methods of construction and analysis of nonlinear error-correcting codes, Doctor of Sciences Diss., Moscow, 1985 (in Russian).